

# Elevating fraud detection: machine learning models with computational intelligence optimization

Cheryl Angelica, Charleen, Antoni Wibowo

Department of Computer Science, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia

## Article Info

### Article history:

Received Jan 4, 2024

Revised Mar 26, 2024

Accepted Apr 12, 2024

### Keywords:

E-commerce

Fraud detection

Genetic algorithm

Light gradient boost machine

Particle swarm optimization

Random forest classifier

X-gradient boost

## ABSTRACT

The amount of crimes committed online has undoubtedly increased as more people use the internet for e-commerce and other financial transactions. Machine learning algorithms have been created to detect payment fraud in online purchasing in order to address the issue. This study performs a thorough comparative examination of different metaheuristic optimizations as hyperparameter tuning methods; these are particle swarm optimization (PSO) and genetic algorithm (GA). They are used to optimize the receiver operating characteristic (ROC) area under the curve (AUC) of the three machine learning algorithms, namely X-gradient boost, random forest classifier, and light gradient boost machine. Since the study's data are unbalanced, the determined metrics were ROC AUC. PSO offers consistent conditions for finding the best solution, according to our experiment. Without the inclusion of population annihilation strategies, PSO can achieve the greatest results in various situations which are different from GA, a consistent condition for finding the best solution, according to our experiment. Without the inclusion of population annihilation strategies, PSO can achieve the greatest results in various situations. The findings indicate that random forest classifier provided the highest ROC AUC value both before and after the hyperparameter tuning process, with a score of 88.69% attained while utilizing PSO.

This is an open access article under the [CC BY-SA](#) license.



## Corresponding Author:

Charleen

Computer Science Department, Master of Computer Science, Bina Nusantara University

Jakarta 11480, Indonesia

Email: charleen@binus.ac.id

## 1. INTRODUCTION

The way society functions has undergone a significant change as a result of technological advancement. The development of technologies has completely changed how people could go about their everyday life and do business. Meeting personal needs is one of the improvements. They only purchase their wants using digital technologies namely online shopping. It has become a popular and practical way to buy goods and services. The number of options for online shopping has significantly increased with the development of companies like Shopee and Tokopedia. However, the increasing usage of the internet for e-commerce and other financial operations has unavoidably increased the number of crimes committed on the internet. The concealment of the complex network may provide a breeding ground for criminals to engage in fraudulent activities [1].

The merchants must deal with a wide variety of fraud and abuse. The account takeover, free-trial abuse, refund fraud, reseller fraud, fake product news, first-party fraud, payment fraud, and many more are examples of fraud [2]. All of these put the merchants' reputation and revenue at risk. Money laundering and the spread of fake news are two examples of fraud that have serious repercussions for society as a whole.

E-commerce fraud is increasing quickly because, unlike in the early years of the internet, today's scammers are well-funded and well-equipped professional rings. The extremely constrained window of time in which acceptance or denial must be made is the main difficulty in fraud detection. The sheer volume of transactions that must be handled at one time is also remarkable. One of the main concerns with e-commerce systems is identifying fraud as soon as possible when the transaction is being performed [3]. In reality, fraudulent and legitimate transactions coexist, which makes it challenging to accurately detect them using basic pattern matching methods [4]. Therefore, this essay focuses on how to spot payment fraud in online shopping.

For the purpose of identifying fraud, a lot of machine learning models have been developed and have been shown to be capable of saving billions of dollars annually. Machine learning is equivalent to having multiple groups of analysts run thousands of queries and evaluate the outcomes to choose the optimal solution. When identifying minor trends or illogical patterns, it is much faster, more scalable, effective, and more accurate than humans. This indicates that the model is able to recognize suspect clients even in the absence of a chargeback.

Hyperparameter tuning is done to improve the machine learning algorithm. Metaheuristic optimization techniques, including particle swarm optimization (PSO) and genetic algorithm (GA), were employed in this research. The basic concept of GA's algorithm is the theory of life evolution. PSO, on the other hand, adopts the theory of colonization from a swarm of feeding creatures [5]. The layout of this essay will be as follows. Related works to this study are included in section 2. The experiment's technique or methodology was explored in section 3. While part 4 will focus on the experiment results analysis. Finally, the conclusion was covered in section 5.

In machine learning, there are various methods for detecting fraud. The effectiveness of fraud detection software is frequently correlated with the efficiency of utilized feature engineering strategies, particularly those that explain client behavior because this information is quite helpful in identifying fraud patterns. Fraud by credit cards is rising at the same rate as the number of customers.

Khatri *et al.* [6] tested the suitability of supervised machine learning models for estimating the likelihood of fraudulent transactions using an unbalanced dataset. They studied some models including random forest, K-nearest neighbor (KNN), decision tree, naive Bayes, and logistic regression. The decision tree model is the most effective one for foretelling these frauds. Although the KNN model is more sensitive than the decision tree, it takes much longer to test the data. Another approach is using machine learning algorithms like the random forest and the Adaboost algorithms. Sailusha *et al.* [7] classify the credit card transactions that are within the dataset as both fraud and non-fraud transactions. In terms of recall, precision, and F1-score, the random forest algorithm surpasses the Adaboost algorithm.

Deep learning has become an important factor among several subfields of finance. Nanduri *et al.* [2], suggested a novel deep learning-based model for rules extraction to identify fraudulent e-commerce transactions. This algorithm surpasses the state-of-the-art in terms of recall, accuracy, area under the receiver operating characteristic (AUROC), and F1-score. The absence of all likely patterns needed to train adequate supervised learning models is commonly the difficulty related with fraud detection. The fact that fraudulent patterns are not just rare but also change over time makes this problem worse. Youssef *et al.* [8] propose a method that, by giving each data point a consistency score, has a great deal of promise for detecting outliers and pure inliers.

Taobao, recognized as one of the largest global e-commerce platforms, boasts a substantial repository of transaction data. The platform employs a sophisticated fraud detection system, extensively evaluated in [9]. The assessment reveals that the proposed attacks exhibit a notable likelihood of circumventing the deployed detector, causing a significant reduction in average precision from nearly 90% to 20%.

Microsoft's new fraud-management system (FMS) explained in [10] can quickly identify and stop new fraud patterns. Advanced real-time dynamic risk feature generation methods and specially created short- and long-term sequential machine learning models are used. The system also optimizes the risk-control module's decisions using dynamic optimization techniques known as prospective control modeling for the highest long-term sales and profits. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

## 2. PROPOSED METHODS

The phases of this research are as follows: i) a preliminary study; ii) data gathering; iii) data pre-processing; and iv) construction of the machine learning models and evaluation. A preliminary study is made in order to fully comprehend the subject of e-commerce fraud, which has been chosen for the research. In the second stage and third stage, the data are gathered and pre-processed to make them appropriate for machine learning models. The various machine learning models will next be compared with our suggested model which is the model with the addition of a meta-heuristic algorithm as their hyperparameter tuning methods.

### 2.1. Preliminary study

This research's initial phase was a preliminary investigation that included literature reviews. The purpose of the literature review was to identify prior research, clarify the theoretical underpinnings, and describe the methodologies employed in this study. Reading sources from journals relating to machine learning models and e-commerce fraud detection, like IEEE and Springer, are searched and reviewed as part of the study of literature.

### 2.2. Data gathering

The IEEE-computational intelligence society (IEEE-CIS) provided the dataset for this study, which was made accessible by Vesta Corporation. The pioneer in secure online payment methods is Vesta Corporation. Identity table and transaction table are two of the data sets that are utilized to create the machine learning model in this study. A transaction table has 590,540 records and 394 feature columns. The transaction table includes information about money transfers as well as other gifting items and services, including product codes, payment card details, client addresses, and transaction amounts. 41 columns of features make up the 144,233 data in the identity database. A digital signature (UA/browser/os/version) and network connection information (IP, ISP, proxy) related to transactions are both stored in the Identity table. They are gathered by the partners in digital security and Vesta's fraud prevention system.

### 2.3. Data pre-processing

Data pre-processing comes next after data collection. Raw data must be modified during data pre-processing in order for it to be used in a machine learning model. It involves data cleaning, data formatting, and data transformation. Dropping the columns with more than 20% of the values missing is the first step in cleaning and formatting the data in this dataset. Then, the category and numerical features are divided and handled differently depending on the kind of each feature. The approach to filling in the missing values for numerical attributes is to use the median. For categorical features, however, the most common value should be used to fill in any missing values. Then, the category and numerical properties are combined once more. For data transformation, by comparing each category level to a predetermined reference level, one-hot encoding is used to convert the categorical features for data transformation [11]. The dataset was then divided into a 20% testing set and an 80% training set. While StandardScaler function is used to standardize the functionality of the input dataset for numerical features.

The dataset has an imbalance class for classification, as shown by the data. Figure 1 representation of the change of dataset's high normal-to-fraudulent ratio illustrates how predictions from the original training set were wildly distorted to balanced dataset. Imbalanced class classification of data is a significant issue for machine learning and hard to improve accuracy. Almost every classification technique will provide substantially higher accuracy for the majority class than for the minority class when working with unbalanced data. An indication of subpar categorization performance is this disparity. In order to resolve this issue, the data was processed with the synthetic minority over-sampling technique (SMOTE) algorithm by oversampling method. Modifying the majority or minority class distribution allows the classification algorithm to treat an imbalanced dataset equally [12]. By resampling the minority class sample, this strategy creates a fresh sample from the minority class to balance the dataset [13].

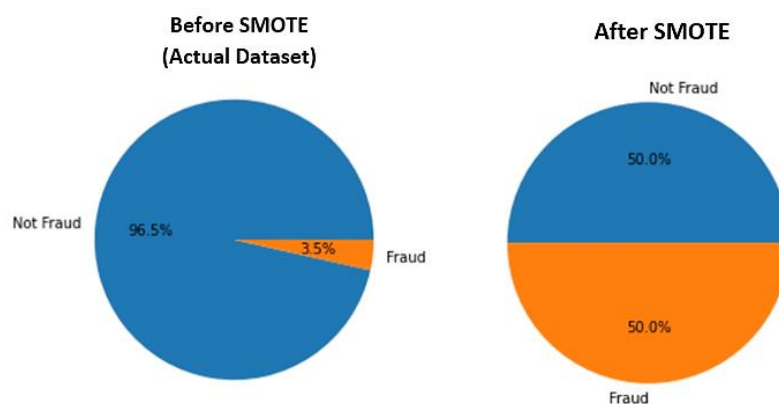


Figure 1. Pie chart of total fraud and not fraud before SMOTE and after SMOTE

## 2.4. Model building and evaluation

Model building and evaluation come next. Figure 2 represents the flow of creating the model. Our approach to determine the optimal machine learning model for spotting ecommerce fraud involves comparing models like random forest, light gradient boost machine, and x-gradient boost. To improve the model's accuracy and receiver operating characteristic (ROC) area under the curve (AUC) value, hyperparameter approaches will be integrated with it. Two hyperparameter optimization approaches, PSO and GA, were applied to enhance the model's quality. John Holland invented GAs, which were influenced by Darwin's theory of evolution. A search algorithm based on the principles of natural selection and genetics is known as a GA [14]. PSO, a stochastic optimization technique, is based on the movement and behavior of swarms. PSO uses the concept of social engagement to resolve a conflict. It makes use of a swarm of particles (called agents) that roam the search space in search of the optimum solution. The positional coordinates in the solution space that correspond to each swarm member's best solution thus far are sought after by each member of the swarm. It is known as a personal best [15]. The optimum hyperparameter approaches for each machine learning model will be determined by comparing them to one another. The evaluation metrics that will be examined are ROC AUC since they can be overly optimistic in situations when there are few samples of the minority class and there is a badly imbalanced classification problem [16].

The evaluation metric focused on in this study is ROC AUC. This is a result of the dataset's imbalanced number of classes, where there are a disproportionately large number of not fraudulent transactions. Accuracy value for data with a uniform distribution is more significant than ROC AUC. However, ROC AUC may be more significant in other circumstances, usually for extremely unbalanced data. ROC-AUC is required to determine how well the model can distinguish between positive and negative classes [17]. While accuracy is simply the proportion of accurate predictions, ROC AUC contrasts the relationship between true positive rate and false positive rate [18].

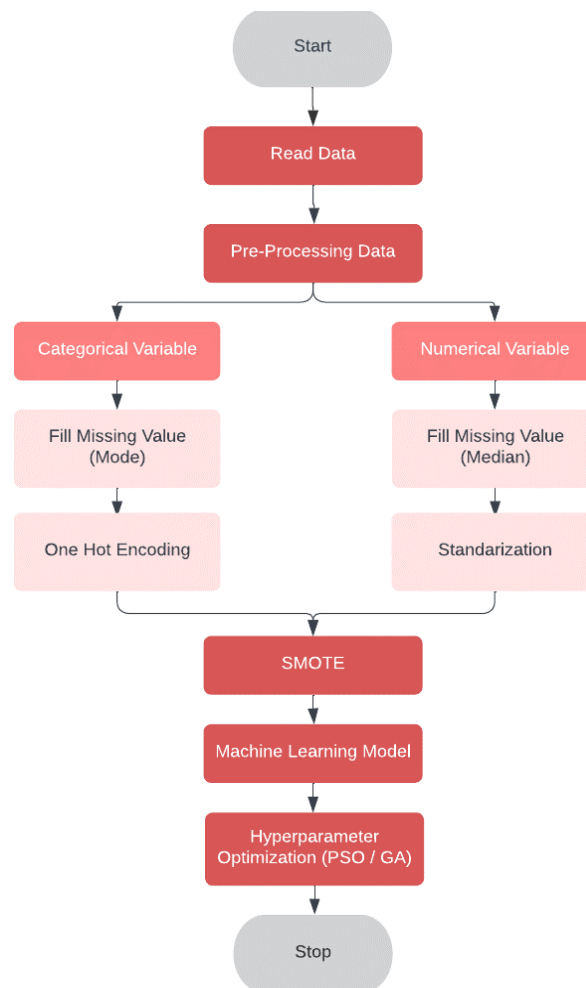


Figure 2. The flow of creating the model

### 3. RESULTS AND DISCUSSION

The findings of the technique previously outlined in section 3 are summarized in Table 1. Based on their ROC AUC value of 88.69%, random forest classifier using PSO as their optimization for hyperparameter tuning produces the greatest results out of all machine learning methods as it is shown in Figure 3. The top machine learning method, however, according to test accuracy, is XG-boost classifier, which has a 97.28% accuracy rate.

Table 1. Summary ROC AUC value and accuracy with and without hyperparameter tuning

Machine learning models		ROC AUC (%)	Train accuracy (%)	Test accuracy (%)
Random forest classifier	Default	88.38	100.00	97.27
	PSO	88.69	98.88	97.22
	GA	87.73	97.06	96.45
XG-Boost classifier	Default	87.53	97.96	97.21
	PSO	87.27	98.43	97.21
	GA	88.33	98.25	97.28
Light gradient boost machine	Default	87.04	96.45	95.72
	PSO	87.64	98.36	97.10
	GA	84.99	97.01	96.76

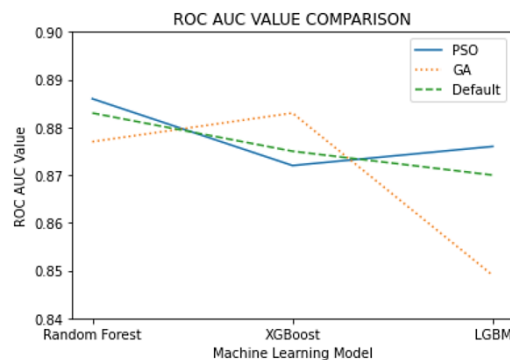


Figure 3. Line graph of ROC AUC value comparison

Here are the findings after each machine learning model's hyperparameters were tuned using PSO and GA techniques, with the ROC AUC calculated appropriately. Table 2 shows the set of best hyperparameters that represent the ideal model configuration achieved according to each strategy and, as a result, the highest accuracy, as well as the corresponding score (ROC AUC) and tuning techniques for hyperparameters. It should be noted that while default settings produce excellent results for certain machine learning models and datasets, this may not be circumstance when the model and/or data source are changed [19]. According to our dataset, setting the default value for the 'learning rate' in the light gradient boost machine using PSO to 0.1 yielded a ROC AUC value of 87.04%; however, setting this hyperparameter to 0.36 yielded a higher ROC AUC value of 87.64%.

According to the confusion matrix in Figure 4, the random forest classifier model with GA as its hyperparameter optimization method correctly predicts the majority of fraud transactions (1,731 data). This result is different from the best ROC AUC value achieved by the random forest classifier model with PSO as its hyperparameter optimization method. This is due to the number of errors in predicting transactions that are not fraud but predicted fraud is much more than the random forest classifier model with PSO as its hyperparameter optimization with a difference of 903 data. The model with the best ROC AUC also managed to predict fraudulent transactions correctly, namely as much as 1,575 data. The error in predicting fraudulent transactions is also small, which is only 775 data. Therefore, the random forest classifier model with PSO as its hyperparameter optimization method gets the best ROC AUC value in predicting fraudulent transactions.

PSO outperforms GAs in some cases, and vice versa, implying that the two methods traverse the problem in different ways. PSO successfully outperforms GA from ROC AUC result using random forest classifier and light gradient boost machine learning model. On the other hand, GA outperforms PSO from ROC AUC value using the XG-Boost classifier model. The findings of GA and PSO implementations in fraud detection prove that the PSO algorithm outperforms the GA algorithm in both accuracy and iteration in determining the best solution. PSO's key advantages over mathematical algorithms and other heuristic optimization methods are its simple concept, straightforward implementation, adaptability to control parameters, and computation time [20].

Table 2. Comparison of the best hyperparameter value in each machine learning model

HP approach	Machine learning model	ROC AUC (%)	Best hyperparameters
PSO	Random forest	88.69	'max_depth':37 'n_estimators':78 'max_features': 37 'min_samples_leaf': 6 'min_samples_split':5
	XG-Boost	87.27	'max_depth': 7 'learning_rate': 0.36 'scale_pos_weight': 2 'gamma': 5 'reg_lambda': 1 'subsample': 0.76 'colsample_bytree': 0.29
	Light gradient boost machine	87.64	'max_depth': -1, 'learning_rate': 0.36 'max_bin': 460 'min_data_in_leaf': 36 'feature_fraction': 0.98
GA	Random forest	87.73	'max_depth':39 'n_estimators':27 'max_features': 8 'min_samples_leaf': 6 'min_samples_split': 7
	XG-Boost	88.33	'max_depth': 10 'learning_rate': 0.11 'scale_pos_weight': 1 'gamma': 4 'reg_lambda': 0 'subsample': 0.91 'colsample_bytree': 0.17
	Light gradient boost machine	84.99	'max_depth': -1 'learning_rate': 0.27 'max_bin': 52 'min_data_in_leaf': 10 'feature_fraction': 0.04

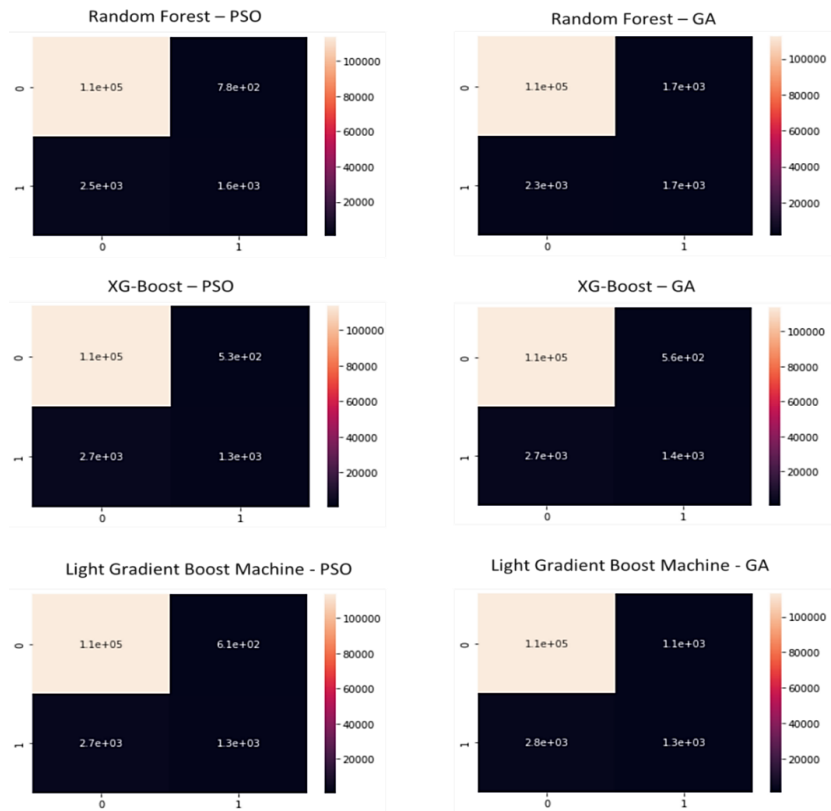


Figure 4. Confusion matrix comparison of machine learning models

Both GA and PSO employ population-based methodologies [21], sharing several characteristics. While GA is recognized for discovering global optimal solutions [22], this research highlights a time advantage for PSO over GA. Computational efficiency favors PSO, as evidenced by its lower computational cost due to more apparent convergence [23] and a simpler parameter setup [24]. Unlike GA, PSO achieves a steady state in locating.

Optimal solutions, eliminating the need for population destruction techniques [5]. Additionally, PSO's simplicity in implementation and ease of parameter tuning makes it particularly appealing for computationally constrained applications. This facilitates faster deployment and requires less fine-tuning compared to GA [25].

#### 4. CONCLUSION

Since today's scammers are professional rings with ample resources and modern equipment, unlike in the early years of the internet, e-commerce fraud is growing swiftly. Many machine learning models have been created for the purpose of recognizing fraud and have proven to be capable of saving billions of dollars per year. It is faster, more scalable, efficient, and precise than humans at spotting insignificant trends or illogical patterns. In this work, a comparison analysis has been conducted to identify the top machine learning algorithms. XG-boost, random forest classifier, and light gradient boost machine models of machine learning are employed. Metaheuristic optimization methods like GA and PSO have been applied as a hyperparameter tuning strategy to improve the quality of machine learning models. Based on our experiment, in contrast to GA, PSO provides a steady condition for finding the best solution. Without the use of population elimination methods, PSO can get the best outcomes in various situations. Unlike GAs, which must eliminate populations after they reach a specific saturation point in order to increase accuracy. PSO also has faster computational time than GA. The findings indicate that random forest classifier provided the highest ROC AUC value both before and after the hyperparameter tuning process, with a score of 88.69% attained while utilizing PSO.

#### REFERENCES




- [1] H. Zhou, G. Sun, S. Fu, L. Wang, J. Hu, and Y. Gao, "Internet financial fraud detection based on a distributed big data approach with Node2vec," *IEEE Access*, vol. 9, pp. 43378–43386, 2021, doi: 10.1109/ACCESS.2021.3062467.
- [2] J. Nanduri, Y. Jia, A. Oka, J. Beaver, and Y. W. Liu, "Microsoft uses machine learning and optimization to reduce e-commerce fraud," *Interfaces*, vol. 50, no. 1, pp. 64–79, 2020, doi: 10.1287/inte.2019.1017.
- [3] J. Shaji and D. Panchal, "Improved fraud detection in e-commerce transactions," *2017 2nd International Conference on Communication Systems, Computing and IT Applications, CSCITA 2017 - Proceedings*, pp. 121–126, 2017, doi: 10.1109/CSCITA.2017.8066537.
- [4] S. B. E. Raj and A. Annie Portia, "Analysis on credit card fraud detection methods," *2011 International Conference on Computer, Communication and Electrical Technology, ICCET 2011*, pp. 152–156, 2011, doi: 10.1109/ICCET.2011.5762457.
- [5] F. D. Wihartiko, H. Wijayanti, and F. Virgantari, "Performance comparison of genetic algorithms and particle swarm optimization for model integer programming bus timetabling problem," *IOP Conference Series: Materials Science and Engineering*, vol. 332, no. 1, 2018, doi: 10.1088/1757-899X/332/1/012020.
- [6] S. Khatri, A. Arora, and A. P. Agrawal, "Supervised machine learning algorithms for credit card fraud detection: A comparison," *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering*, pp. 680–683, 2020, doi: 10.1109/Confluence47617.2020.9057851.
- [7] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. R. Rao, "Credit card fraud detection using machine learning," *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020*, pp. 1264–1270, 2020, doi: 10.1109/ICICCS48265.2020.9121114.
- [8] B. Youssef, F. Bouchra, and O. Brahim, "Rules extraction and deep learning for e-commerce fraud detection," *Colloquium in Information Science and Technology, CIST*, vol. 2020, pp. 145–150, 2020, doi: 10.1109/CiSt49399.2021.9357066.
- [9] U. Porwal and S. Mukund, "Credit card fraud detection in E-commerce," *Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019*, pp. 280–287, 2019, doi: 10.1109/TrustCom/BigDataSE.2019.00045.
- [10] Q. Guo *et al.*, "Securing the deep fraud detector in large-scale e-commerce platform via adversarial machine learning approach," *The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019*, pp. 616–626, 2019, doi: 10.1145/3308558.3313533.
- [11] K. Potdar, T. S. Pardawala, and C. D. Pai, "A comparative study of categorical variable encoding techniques for neural network classifiers," *International Journal of Computer Applications*, vol. 175, no. 4, pp. 7–9, 2017, doi: 10.5120/ijca2017915495.
- [12] H. Liu, M. Zhou, and Q. Liu, "An embedded feature selection method for imbalanced data classification," *IEEE/CAA Journal of Automatica Sinica*, vol. 6, no. 3, pp. 703–715, 2019, doi: 10.1109/JAS.2019.1911447.
- [13] A. Fernández, S. García, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary," *Journal of Artificial Intelligence Research*, vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.
- [14] T. Friedrich, T. Kötzing, M. S. Krejca, and A. M. Sutton, "The compact genetic algorithm is efficient under extreme Gaussian noise," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 3, pp. 477–490, 2017, doi: 10.1109/TEVC.2016.2613739.
- [15] H. T. Rauf, U. Shoaib, M. I. Lali, M. Alhaisoni, M. N. Irfan, and M. A. Khan, "Particle swarm optimization with probability sequence for global optimization," *IEEE Access*, vol. 8, pp. 110535–110549, 2020, doi: 10.1109/ACCESS.2020.3002725.
- [16] A. J. Bowers and X. Zhou, "Receiver operating characteristic (ROC) area under the curve (AUC): a diagnostic measure for






- evaluating the accuracy of predictors of education outcomes,” *Journal of Education for Students Placed at Risk*, vol. 24, no. 1, pp. 20–46, 2019, doi: 10.1080/10824669.2018.1523734.
- [17] A. M. Carrington *et al.*, “Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329–341, 2023, doi: 10.1109/TPAMI.2022.3145392.
- [18] W. Deng, Z. Huang, J. Zhang, and J. Xu, “A data mining-based system for transaction fraud detection,” *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*, pp. 542–545, 2021, doi: 10.1109/ICCECE51280.2021.9342376.
- [19] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, “Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis,” *Informatics*, vol. 8, no. 4, 2021, doi: 10.3390/informatics8040079.
- [20] M. Issa, A. E. Hassanien, D. Oliva, A. Helmi, I. Ziedan, and A. Alzohairy, “ASCA-PSO: Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment,” *Expert Systems with Applications*, vol. 99, pp. 56–70, 2018, doi: 10.1016/j.eswa.2018.01.019.
- [21] S. T. Li, B. Zhang, S. J. Xu, and Y. H. Zhong, “Back-analysis of pavement thickness based on PSO-GA hybrid algorithms,” *IOP Conference Series: Earth and Environmental Science*, vol. 252, no. 5, 2019, doi: 10.1088/1755-1315/252/5/052066.
- [22] A. Yazdanpanah, A. Rezaei, H. Mahdiyar, and A. Kalantariasl, “Development of an efficient hybrid ga-pso approach applicable for well placement optimization,” *Advances in Geo-Energy Research*, vol. 3, no. 4, pp. 365–374, 2019, doi: 10.26804/ager.2019.04.03.
- [23] Y. Ding, W. Zhang, L. Yu, and K. Lu, “The accuracy and efficiency of GA and PSO optimization schemes on estimating reaction kinetic parameters of biomass pyrolysis,” *Energy*, vol. 176, pp. 582–588, 2019, doi: 10.1016/j.energy.2019.04.030.
- [24] S. M. Almufti, A. Y. Zebari, and H. K. Omer, “A comparative study of particle swarm optimization and genetic algorithm,” *Journal of Advanced Computer Science & Technology*, vol. 8, no. 2, pp. 40–45, 2019, doi: 10.14419/jacst.v8i2.29401.
- [25] A. A. Karim, N. A. M. Isa, and W. H. Lim, “Modified particle swarm optimization with effective guides,” *IEEE Access*, vol. 8, pp. 188699–188725, 2020, doi: 10.1109/ACCESS.2020.3030950.

## BIOGRAPHIES OF AUTHORS






**Cheryl Angelica**    is a master's student at the School of Computer Science, Bina Nusantara University, Indonesia. She entered college in 2019 and completed undergraduate studies majoring in Information Technology (IT) in 2022. She then immediately takes a master's degree with the same major and is currently a last year student. Her research interests are primarily in the area of artificial intelligence application, especially in machine learning, deep learning, and data science. She can be contacted at email: [cheryl.angelica@binus.ac.id](mailto:cheryl.angelica@binus.ac.id).



**Charleen**    is a dedicated college student in the Bina Nusantara University Graduate Program (BGP). Pursuing a degree in Computer Science, she has actively contributed to this paper through writing this paper. Showcasing a strong foundation in computer vision. With a passion for research and a well-rounded approach to education, she aspires to make significant future contributions to the field of computer science. She is currently in her 9th semester of her college year at Bina Nusantara University, Jakarta 11480, Indonesia. She can be contacted at email: [charleen@binus.ac.id](mailto:charleen@binus.ac.id).



**Antoni Wibowo**    (M'12) received my first degree of Applied Mathematics in 1995 and master degree of Computer Science in 2000. In 2003, he awarded a Japanese Government Scholarship (Monbukagakusho) to attend Master and Ph.D. programs at Systems and Information Engineering in University of Tsukuba-Japan. He completed the second master's degree in 2006 and Ph.D. degree in 2009, respectively. His Ph.D. research focused on machine learning, operations research, multivariate statistical analysis, and mathematical programming, especially in developing nonlinear robust regressions using statistical learning theory. He has worked from 1997 to 2010 as a researcher in the Agency for the Assessment and Application of Technology–Indonesia. From April 2010–September 2014, he worked as a senior lecturer in the Department of Computer Science–Faculty of Computing, and a researcher in the Operation Business Intelligence (OBI) Research Group, Universiti Teknologi Malaysia (UTM)–Malaysia. From October 2014–October 2016, he was an Associate Professor at Department of Decision Sciences, School of Quantitative Sciences in Universiti Utara Malaysia (UUM). He is currently working at Binus Graduate Program (Master in Computer Science) in Bina Nusantara University-Indonesia as a specialist lecturer and continues his research activities. He can be contacted at email: [anwibowo@binus.edu](mailto:anwibowo@binus.edu).